

What Is Claimed Is:

1. A method of predicting cache performance, comprising:
storing data references applied to an operational data cache in a data
5 processing environment;
applying said data references to a cache simulator configured to
simultaneously simulate a plurality of caches of different sizes, wherein said
cache simulator comprises multiple segments and each said simulated cache
comprises one or more of said segments; and
10 generating for each of said plurality of simulated caches an estimate of
performance based on said simulation;
wherein each application of one of said data references to said cache
simulator causes either a hit in one of said segments or a miss of every said
segment.
15
2. The method of claim 1, further comprising:
dynamically adjusting the size of said operational data cache to match the
size of one of said simulated caches.
- 20 3. The method of claim 1, wherein said data processing environment
is a database management system and said operational data cache is a buffer cache
configured to cache data in said database management system.
4. The method of claim 3, wherein said storing data references
25 comprises:
receiving a first data reference at said operational data cache;
storing said first data reference in a trace buffer for use in said simulation

of said plurality of caches if said trace buffer is not full; and
discarding said first data reference if said trace buffer is full.

5 5. The method of claim 1, wherein said cache simulator comprises a
list of simulated buffers, and each said simulated buffer is configured to store:
an identifier of a data item; and
an identifier of said segment of said cache simulator in which said
simulated buffer is located.

10 6. The method of claim 5, wherein said applying comprises:
retrieving a first stored data reference;
identifying a first data item referenced in said first data reference; and
searching said cache simulator for a buffer in which an identifier of said
first data item is stored.

15 7. The method of claim 6, further comprising:
if a first buffer is found in said cache simulator that stores an identifier of
said first data item:
incrementing a hit counter for said segment in which said first
20 buffer is found;
moving said first buffer to a head of said cache simulator; and
updating said stored segment identifiers of said buffers as
necessary.

25 8. The method of claim 6, further comprising:
if no buffer is found in said cache simulator that stores an identifier of said
first data item:

incrementing an absolute miss counter;
storing an identifier in a first buffer in said cache simulator;
moving said first buffer to a head of said cache simulator; and
updating said stored segment identifiers of said buffers as
5 necessary.

9. The method of claim 1, wherein said generating comprises:
for a first simulated cache in said plurality of simulated caches:

calculating the number of hits in all segments of said cache
10 simulator that do not comprise part of said first simulated cache; and
adding the number of said misses of all of said segments to said
calculated number of hits to produce an initial estimated miss rate for said
first simulated cache.

15 10. The method of claim 9, wherein said generating further comprises:
calculating a correction factor to apply to said initial estimated miss rate
for said first simulated cache.

20 11. The method of claim 10, wherein said correction factor comprises
the ratio of misses incurred during application of said data references to said
operational data cache to an initial estimated miss rate for a second simulated
cache in said plurality of simulated caches;

wherein the number of buffers in said second simulated cache matches the
number of buffers in said operational data cache.

25

12. The method of claim 10, wherein said generating further
comprises:

multiplying said initial estimated miss rate for said first simulated cache by said correction factor to yield a predicted miss rate for said first simulated cache.

- 5 13. A method of simulating the performance of multiple caches of different sizes, comprising:
- storing references to data items received during operation of a database management system;
- maintaining a multi-segmented cache simulator comprising simulated
- 10 buffers configured to store data identifiers, wherein each of said multiple caches comprises a set of said segments different from the other caches;
- for each of said stored references:
- searching said cache simulator for a first simulated buffer storing an identifier of said referenced data item;
- 15 if said first simulated buffer is found in said segmented memory:
- incrementing a hit counter for said segment in which said first simulated buffer is located; and
- moving said simulated cached reference to the head of said segmented memory;
- 20 if no simulated buffer storing an identifier of said referenced data item is found in said cache simulator:
- incrementing a miss counter;
- storing an identifier of said data item in a second simulated buffer; and
- 25 storing said second simulated buffer at a head of said cache simulator; and
- generating an estimate of the performance of each of said multiple caches.

14. The method of claim 13, wherein said estimated performance comprises a number of misses, wherein said generating comprises:

5 for a first cache of said multiple caches, calculating the number of said referenced data items for which identifiers were not found in the cache simulator segments comprising said first cache.

15. The method of claim 14, further comprising:

calculating a predicted number of misses for said first cache by
10 multiplying said estimated performance for said first cache by a correction factor.

16. The method of claim 15, wherein:

said data item references are received at an operational cache of said database management system and result in a known number of misses in said
15 operational cache;

a second cache of said multiple caches comprises a number of simulated buffers equivalent to the number of buffers comprising said operational cache; and

said correction factor is equal to the ratio of said known number of misses
20 to said estimated performance for said second cache.

17. The method of claim 16, further comprising:

dynamically altering the size of said operational cache to the size of said first cache, wherein said predicted number of misses for said first cache is less
25 than said known misses in said operational cache.

18. The method of claim 13, wherein said storing comprises:

receiving a first reference to a data item at said database management system;

if a trace memory configured to store said first reference is full, discarding said first reference; and

5 otherwise, storing said first reference in said trace memory without locking said trace memory.

19. The method of claim 13, wherein said data item references are received at an operational cache of said database management system, further comprising:

identifying the number of operational cache misses resulting from application of said data item references to said operational cache; and

dynamically altering the size of said operational cache to the size of one of said multiple caches having a superior estimated performance.

15

20. A computer readable storage medium storing instructions that, when executed by a computer, cause the computer to perform a method of predicting cache performance, the method comprising:

storing data references applied to an operational data cache in a data processing environment;

applying said data references to a cache simulator configured to simultaneously simulate a plurality of caches of different sizes, wherein said cache simulator comprises multiple segments and each said simulated cache comprises one or more of said segments; and

25 generating for each of said plurality of simulated caches an estimate of performance based on said simulation;

wherein each application of one of said data references to said cache

simulator causes either a hit in one of said segments or a miss of every said segment.

21. A computer readable storage medium containing a data structure
5 configured for simulating multiple caches, of different sizes, for a set of data item references applied to the multiple caches, the data structure comprising:
a list of simulated buffers, wherein each buffer is configured to store:
an identifier of a data item; and
an identifier of a portion of said list of buffers in which said buffer
10 is located;
a miss counter configured to increment each time the data item of an applied data reference does not correspond to any of said data item identifiers of said simulated buffers; and
for each said portion, a hit counter configured to increment each time a
15 buffer in said portion is found to store an identifier of the data item of an applied data reference.

22. The computer readable storage medium of claim 21, wherein said data structure further comprises:
20 for each said portion of said list, an identifier of a head of said portion.

23. A system for simulating the performance of multiple caches, comprising:
a reference memory configured to store data references;
25 a segmented memory of simulated buffers, wherein each of said multiple simulated caches comprises one or more of said memory segments; and
an engine configured to apply said data references to said segmented

memory.

24. The system of claim 23, wherein each of said simulated buffers is configured to store:

- 5 an identifier of a data item; and
 a segment identifier configured to identify which of said memory segments includes said simulated buffer.

25. The system of claim 23, said multiple caches consisting of N
10 caches and said segmented memory consisting of N segments, wherein:
 simulated cache 1 consists of memory segment 1; and
 for simulated caches $M = 2$ to N, each said simulated cache M consists of
 said memory segments 1 to M.

- 15 26. The system of claim 23, further comprising an operational cache,
 wherein said data references are references received during operation of said
 operational cache.

27. The system of claim 24, wherein said data references are references
20 to data items received in a database management system.